https://doi.org/10.36375/prepare_u.iei.a128

Statistical modelling for rainfall time series analysis: Khurdha district of Odisha, India

Ankita Bohidar[1] , Dr. Anil Kumar Kar[2] , Prof. Pradeep Kumar Das[3]

[1] Research Scholar, Civil Engineering Department, VSSUT, Burla-768018, Sambalpur, Odisha
[2] Associate Professor, Civil Engineering Department, VSSUT, Burla-768018, Sambalpur, Odisha
[3] Registrar, NIT Rourkela,Rourkela-769008, Odisha.

Email: ankita_bohidar@rediffmail.com, anilkarhy@gmail.com

Abstract:

The rainfall at Odisha state is monsoon driven. The capital city of Odisha state is Bhubaneswar which lies at Khurdha district. In this study, a statistical modeling is done for the monsoon rainfall of this district including frequency analysis of the monsoon rainfalls using L-moment techniques. The randomness of the data is determined from Anderson Correllogram test and then the existence of probable trend is determined using non-parametric test like Mann Kendall. All the datasets are found to be random but the rainfall during August shows a rising trend at all 1, 5 and 10% significance level. Also in month July it rises 5 and 10% Significance level. The forecasting of the monthly rainfall is made through an Auto Regressive Moving Average (ARMA) model. The ARMA (1,1) combination hold good for months of June, July and September, August and October ARMA (1,2) ARMA (3,3) respectively show better result. Akaike Information Criteria (AIC) has been used for evaluating the performance of ARMA models. The study shows the statistical application on climate data and the results are advisory and indispensable for making useful and reliable decisions in hydrological forecasting and planning.

Key words: Trend, Mann Kendall, Anderson Correllogram, L-moment, Khurdha

Introduction: Statistical analyses of hydrological time series play a vital role in water resources studies. The statistical results reflect the inherent characteristics of a data set. Non parametric tests in detection of trend are a regular practice. Particularly in a climate change scenario detecting trend in a hydro-meteorological element like rainfall has a different meaning. As rainfall is the prime factor for all round growth of a locality estimating the variation and frequency of rainfall is important in context of agriculture, design and construction of storage structures as well as for flood hazards. In this study rainfall data of a coastal district named Puri of India is first analyzed for its randomness using Anderson Correllogram test then for existence of trend by using non-parametric Mann Kendall test. The rainfall data is modeled through ARMA model in order to detect the best possible combination of future forecasting.

Gangyan et al. (2002) examined the temporal and spatial sediment load characteristics and used statistical tests such Turning point test, Kendall Rank Correlation test, Anderson correlogram test for identifying the existence of randomness and trend. The periodicity in the sediment load data was analysed by Harmonic Analysis and stochastic component was modeled by auto regressive model. Gao et al. (2002) have applied stochastic hydrology methods to analyze the characteristics of annual inflow evolution of Miyun reservoir. Jain and Kumar (2012) have detected trends in rainfall, rainy days and temperature is being analysed through Sen's non-parametric estimator of slope and statistical significance by Mann Kendall test. Sen test and Mann Kendall test are being applied in many studies as in Vousaghi et al. (2013), Kundu et al. (2014). Chhabra et al. (2014) have applied non-parametric tests to identify trend utilizing 1'x1' gridded rainfall data over North India. In upper Mahanadi, Jaiswal et al. (2014) have made trend assessment for extreme rainfall indices with reference to climate change. They have taken six raingauges for analysis and found no significant trend in any raingauge station while rising trend is seen in very heavy precipitation days in most of the stations. Sethy et al. (2015) performed a trend analysis for precipitation and inflows time series for Salia river basin of Odisha, India which is draining to Chilka lake using the Mann–Kendall test. Dawson et al. (2015) have studied about trends in water quality and quantity for 11 major reservoirs of the Brazos and Colorado river basins in the southern Great Plains. The study of Manee et al. (2015) applied the Mann-Kendall (MK) statistical trend test to analyze increasing,

https://doi.org/10.36375/prepare_u.iei.a128

decreasing or trendless characteristics of precipitation, temperature, inflow to dam reservoirs, release from dam reservoirs, and storage volume in dam reservoir in Thailand from historical operation recorded data. Jaiswal et al. (2015) have made an assessment of change detection and trend on monthly, seasonal and annual historical series of different climatic variables of Raipur, the capital of Chhatisgarh. One of the most useful descriptive tools in time series analysis is to generate the correlogram plot which is simple a plot of the serial correlations rk versus the lag k for k = 0, 1, . . . , M, where M is usually much less than the sample size n. If we have a random series of observations that are independent of one another, then the population serial correlations will all be zero. However, in this case, we would not expect the sample serial correlations to be exactly zero since they are all defined in terms y etc. However, if we do have a random series, the serial correlations should be close to zero in value on average. One can show that for a random series,

$$E[rk] \approx -1\,(n-1) \text{ and}$$
$$Var\,(rk) \approx 1/n$$

In addition, if the sample size if fairly large (say n ≥ 40), then r_k is approximately normally distributed (Kendall et al 1983). The approximate normality of the r_k can aid in determining if a sample serial correlation is significantly non-zero, for instance by examining if r_k falls within the confidence limits $-1/(n-1) \pm 1.96/\sqrt{n}$.

To identify trend in climatic variables with reference to climate change, the Mann-Kendall test has been employed by a number of researches with temperature, precipitation and stream flow data series (Burn, 1994, Douglas et. al 2002, Yue and Hashimo 2003, Burn et al. 2004, Lindst$o$rm and Bergstrom, 2004). It is a common practice to use a non parametric test to detect a trend in a time series. This test, being a function of the ranks of the observations rather than their actual values, is not affected by the actual distribution of the data and is less sensitive to outliers. On the other hand, parametric trend tests, although more powerful, require the data to be normally distributed and are more sensitive to outliers. The Mann–Kendall test is therefore more suitable for detecting trends in hydrological time series, which are usually skewed and may be contaminated with outliers. This test has been extensively used with environmental time series (Hipel and McLeod, 2005).

The Mann-Kendall trend test is based on the correlation between the ranks of a time series and their time order. For the statistics S is calculated as equation (1). This statistic represents the number of positive differences minus the number of negative differences for all the differences considered as

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \qquad (1)$$

where n is the number of total data points, x_i and x_j are the data values in time series i and j $(j > i)$, respectively, and $sgn(x_j - x_i)$ is the sign function as:

$$sgn(x_j - x_i) = \begin{Bmatrix} +1, & if\ (x_j - x_i) > 0 \\ 0, & if\ (x_j - x_i) = 0 \\ -1, & if\ (x_j - x_i) < 0 \end{Bmatrix} \qquad (2)$$

The variance of Mann- Kendall test is calculated by equation (3) as

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5)}{18} \qquad (3)$$

where n is the number of total data points, m is the number of tied groups. The tied group means a simple data having a same value. The t_i indicates the number of ties of extent i. In case of the sample size $n > 10$, the standard normal test statistic Z_s is estimated by equation (4) as

$$Z_s = \begin{bmatrix} \frac{S-1}{\sqrt{Var(S)}}, \ if\ S > 0 \\ 0\ if\ S = 0 \quad (4) \\ \frac{S+1}{\sqrt{Var(S)}}, if\ S < 0 \end{bmatrix} \qquad (4)$$

The positive values of Z_s show increasing trends while negative values represent falling trends. As 5 % significance level is taken standard for this study, the null hypothesis of no trend is rejected if $|Z_s| > 1.96$.

ARMA model developed by Box and Jenkins (1970) provides one of the basic tools in time series modeling. The modeling and forecasting procedures in identifying patterns in time series data involve knowledge about the mathematical model of the process.

First, for a series x_t , we can model that the level of its current observations depends on the level of its lagged observations The AR (1) (autoregressive of order one) can be written as:

$$x_t = \phi x_{t-1} + \in_t$$

Where

$$\in_t \approx WN(0, \sigma_t^2)$$

Similarly, AR(p) (autoregressive of order p) can be written as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots\ldots + \phi_p x_{t-p} + \in_t$$

https://doi.org/10.36375/prepare_u.iei.a128

The MA (1) (moving average of order one) and MA (q) (moving average of order q) can be written as

$$x_t = \in_t + \theta \in_{t-1}$$

and

$$x_t = \in_t + \theta_1 \in_{t-1} + \ldots\ldots + \theta_q \in_{t-q}$$

If we combine these two models, we get a general ARMA (p, q) model,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots\ldots + \phi_p x_{t-p} + \in_t + \theta_1 \in_{t-1} + \ldots$$

The performance of calibration and validation is highly dependent on the structure of the model and the parsimony. The most prominent, and still widely used, criterion is the Akaike Information Criterion (AIC), proposed by Akaike (1974). Akaike Information Criteria (AIC) is a widely used measure of a statistical model. It basically quantifies 1) the goodness of fit, and 2) the simplicity/parsimony, of the model into a single statistic.

$$AIC = m \ln(RMSE) + 2n$$

Where, m is the number of input–output patterns used for training, n is the number of parameters to be identified and RMSE is the root-mean-square error between the network output and target. The performance measures generally improve as more parameters are added to the model, but the AIC statistics penalize the model for having more parameters and, therefore, tend to result in more parsimonious models.

Results and Discussion:

In this study the monthly rainfall data for monsoon season of Khurda district (Fig.1) is taken into consideration for a period of 120 years.

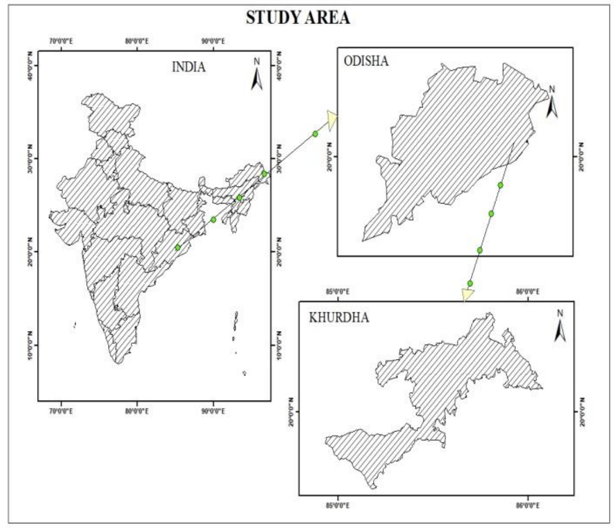

Fig.1 Study area, Khurdha district

The data is initially tested for randomness using Anderson Correlogram test for 1%, 5% and 10% significance level. The testing criteria Z value is obtained for these significance levels (Table 1). It is found that in most of the case the data is random other than in the month of August at 10 % significance level.

Table 1: Randomness check using Anderson Correlogram test

Month	R1	Z	Significance Level		
			1%	5%	10%
June	-0.037	-0.310	Random	Random	Random
July	-0.07	-0.674	Random	Random	Random
August	-0.161	-1.676	Random	Random	Not Random
September	-0.085	-0.844	Random	Random	Random
October	0.09	1.078	Random	Random	Random

To know the statistical values of the rainfall series, all the statistical calculations like mean, standard deviation, coefficient of variation, Coefficient of Skewness, Kurtosis are calculated (Table 2). The month August is showing the highest average rainfall as 250.15 mm whereas in June it is

https://doi.org/10.36375/prepare_u.iei.a128

166.47 mm. it is also revealed that in every month the highest rainfall is above 500 mm except June.

Table 2 Statistics of monthly rainfall series

	Jun	Jul	Aug	Sep	Oct
Avg	166.47	243.50	250.15	218.62	198.20
Max	398.49	614.80	559.20	513.10	663.90
Min	33.17	55.42	96.30	19.86	16.87
Cs	0.85	1.00	0.93	0.52	0.99
Ck	0.62	1.98	3.39	0.65	1.19

The existence of trend is also tested using non-parametric Kendall Rank Test using the significance levels of 1%, 5% and 10% (Table 3). The no trend has been detected in the month of June, September and October where as rising trend in July and August Rising trend of July and August is showing increase in frequency of depressions/ cyclones, because Khurdha is being a coastal district often remains exposed to cyclonic rainfalls The district also remains on the cyclone tract that enters to the state of Odisha. However the falling trend of rainfall is not seen during any of the monsoon season.

Table 3: Trend check using Mann Kendall test

Month	P	Z	Significance Level		
			1%	5%	10%
June	3514	-0.25404	No Trend	No Trend	No Trend
July	4524	3.10295	Rising Trend	Rising Trend	Rising Trend
August	4175	2.74457	Rising Trend	Rising Trend	Rising Trend
September	3629	0.26765	No Trend	No Trend	No Trend
October	3765	0.88461	No Trend	No Trend	No Trend

The ARMA model is tried for modeling of the time series in monthly basis with different Auto Regressive and Moving Average combinations. In both the cases the trial is done from 0, 1, 2, 3 for p, q values at different combinations. The outputs are recorded according to AIC values fixed as the performance criteria (Table 4.1 to 4.12).

Table 4.1 AIC values for the month of June

q in MA(q)	p in AR(p)			
	0	1	2	3
0		1457	1441	1427
1	1526	1395	1397	1397
2	1504	1397	1399	1401
3	1489	1397	1401	1403

Table 4.2 AIC values for the month of July

q in MA(q)	p in AR(p)			
	0	1	2	3
0		1503	1482	1467
1	1601	1437	1439	1439
2	1568	1438	1441	1443
3	1545	1439	1443	1442

Table 4.3 AIC values for the month of August

q in MA(q)	p in AR(p)			
	0	1	2	3
0		637.2	622.4	602.7
1	696.4	603.4	602.7	597.5
2	681.6	589.2	606.6	606.7
3	668.9	591.2	593.2	588.9

Table 4.4 AIC values for the month of September

q in MA(q)	p in AR(p)			
	0	1	2	3
0		1478	1458	1449
1	1575	1412	1414	1415
2	1541	1414	1414	1416
3	1517	1415	1418	1417

Table 4.5 AIC values for the month of October

q in MA(q)	p in AR(p)			
	0	1	2	3
0		1555	1540	1524
1	1599	1499	1501	1502
2	1585	1501	1501	1504
3	1574	1502	1503	1495

Taking the AIC value as performance criteria monthwise best combinations are retrieved from Table 4.1 to 4.5. The ARMA coefficients are also derived for the said best combinations. For the future forecasting the same coefficients may be utilized for finding the respective monthly values.

Table 5, AR, MA coefficients for best combinations in different months

Month	Best combinations (p,q)	ARMA coefficients		AIC values
		p	q	
June	1,1	-1	-0.99956	1395
July	1,1	-0.99974	-0.93586	1437

https://doi.org/10.36375/prepare_u.iei.a128

August	1,2	-0.99977	-1.0832	589.2
		0	0.16457	
September	1,1	-1	-0.99972	1412
October	3,3	0.74082	0.96641	1495
		-0.9839	-0.96369	
		-0.75692	-0.99712	

Conclusion: The statistical modeling is applied on the rainfall data of Khurdha district. The statistical parameters are obtained over the study periods. The rainfall values are found to be random. The non-parametric trends are obtained which shows the possibility of rising trend during the month of July and August whereas a no trend scenario is visible during rest of the monsoon period. In none of the cases falling trend is not seen. This indicates the occurrence of number of cyclonic and depression led rainfall is increasing as the district lies in the coastal part and close to the route of cyclones. Further the ARMA model is applied in order to find the coefficients for the forecasting of the rainfall series. The monthwise best p, q combinations are determined and basing on the AIC values the coefficients are obtained for the ARMA model.

References:

[1] Akaike H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

[2] Box,G.E.P. & Jenkins, G.M. (1970) Time Series Analysis, Forecasting and Control, Holden Day, CA, Sanfrancisco.

[3] Burn, D. H. (1994) Hydrologic effects of climatic change in west-central Canada, Journal of Hydrology, Vol.160 (1), pp.53–70.

[4] Burn, D. H., Cunderlik, . J. M. and Pietroniro, A. (2004) Hydrological trends and variability in the Liard River basin, Hydrological Sciences Journal, Vol. 49(1), pp.53–67.

[5] Caloiero, T., Coscarelli, R. and Ferrari, E. (2018) Analysis of monthly trend in Calabria (Southern Italy) through the application of statistical and graphical techniques, Proceedings, 2,629, DOI: 10.3390/Proceedings2110629.

[6] Dawson, D., VanLandegham, M.M., Asquith, W.H., Patino, R. (2015) Long term trends on reservoir water quality and quantity in two major river basins of Southern great plains., Journal of Lake and Reservoir Management, Vol. 31(3) pp. 254-279.

[7] Douglas,E.M., Vogel,R.M., and Kroll, C.N (2002) Impact of streamflow persistence on hydrologic design, Journal of Hydrologic Engg, Vol.7(3), pp.220–228.

[8] Gangyan, Z, Goel, N.K and Bhatt, V.K (2002) Stochastic modeling of the sediment load of the upper Yangtze river (China), Hydrological Sciences Journal, 47 (S), pp.S-93-S-105.

[9] Gupta, K.K., Kar, A.K., Jena, J. and Jena, D.R. (2017) Variation in rainfall trend at upstream - A threat towards filling schedule of Hirakud Reservoir, India, International Journal of Water, 11(4), 395-407.

[10] Hipel, K and McLeod, A (2005) Time Series Modelling of Water Resources and Environmental Systems, http://www.stats.uwo.ca/faculty/aim/1994Book.

[11] Jaiswal, R.K., Tiwari, H.L. and Lohani, A.K. (2014) Trend assessment for extreme rainfall indices in the upper Mahanadi with reference to climate change, International Journal of Scientific Engineering and Technology (IJSET), 18–20 December, pp.93–99.

[12] Krishnan, G., Chandniha, S.K. and Lohani, A.K. (2015) ‘Rainfall trend analysis of Punjab India using statistical non-parametric test’, Current World Environment, Vol. 10, No. 3, pp.792–800.

[13] Manee, D., Y.Tachikawa, Y., and K.Yorozu, K, (2015) Analysis of Hydrologic Variables Changes related to Large Scale Reservoir Operation by Using Mann-Kendall Statistical Tests in Thailand, THA 2015 International Conference on “Climate Change and Water & Environmental Management in Monsoon Asia” 28-30 January 2015, Bangkok, Thailand.

[14] Portela, M.M., Espinosa, L.A. and Zelenakova, M. (2020) long term rainfall trends and their variability in mainland Portugal in the last 106 years, Climate, 8, 146., DOI: 10.3390/ Cli8120146.

[15] Salami A. W. (2011) Assessment of the impact of climate change on water resources of Jebba hydropower reservoir using trend and reduction pattern methods, Paper presented during the 10th Anniversary Conference of Fulbright held at Conference Centre, University of Ibadan, Ibadan. 12th - 15th, Sept. 2011.

[16] Sethy, R., Pandey, B.K., Krishnan, R. Khare, D. and Nayak, P.C. (2015) Performance evaluation and hydrological trend detection of a reservoir under climate change condition, Journal of Model Earth Syst. Environ (2015) 1:33, pp. 2-10.

[17] Yue, S and Hashino, M (2003) “Long term trends of annual and monthly precipitation in Japan,” Journal of American Water Resources Assoc. Vol. 39(3), pp.587–596.

[18] Zhang, G.P (2007) “A neural network ensemble method with jittered training data for time series forecasting”, Information Sciences 177, pp: 5329–5346.

[19] Zhang,G.P. (2003) Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50 (2003), pp.. 159–175.

[20]